

3. Ytre-Arne B. Introduction: Media Use and Everyday Life in Digital Societies. Media Use in Digital Everyday Life. Emerald Publishing Limited. Leeds. 2023. P. 1-16.

Євгенія Бірюкова, здобувачка першого
(бакалаврського) рівня вищої освіти
навчально-наукового інституту
психології та соціальних наук
Міжрегіональної академії управління
персоналом.
Науковий керівник – кандидат
філологічних наук, професор кафедри
журналістики **Ганна Холод**

ІНСТРУМЕНТИ ВИЯВЛЕННЯ ДІПФЕЙКУ НА ПРИКЛАДІ СУЧАСНИХ МАТЕРІАЛІВ

У цій статті досліджено використання дівфейків у сучасних матеріалах, апробовано конкретні інструменти для виявлення дівфейку в трьох матеріалах. Дослідження сприяє підвищенню обізнаності про дівфейки та їхній вплив на сучасне суспільство.

Ключові слова: дівфейк, медіа, журналістика, соціальні мережі, ЗМІ, дані, аналіз, штучний інтелект.

У ХХІ столітті інформаційні процеси, що ґрунтуються на комп'ютерній обробці великих обсягів даних, стали важливими в усіх сферах суспільного життя. «Політично мотивована дезінформація, безумовно, не є новим явищем, але технологічний прогрес робить створення та розповсюдження маніпулятивного контенту набагато простішим та більш ефективним, ніж у часи до розповсюдження інтернету» [1, с. 9]. Новітні методи пропаганди, залякування, спотворення інформації стали справжньою небезпекою для світу, особливо для України. Технологія «дівфейк» – це найбільш небезпечна зброя в руках любителів спотворювати інформацію й видавати брехню за правду. Ця технологія використовує штучний інтелект, щоб імітувати обличчя людини й робити будь-які фото та відеоматеріали, які забажає недобросовісний

користувач. Багато людей стурбовані тим, що вебсайти можуть поширювати дезінформацію, видаючи їх за «справжні» новини, тому дослідження інструментів виявлення дідфейку, на нашу думку, є актуальним.

Об'єкт дослідження – дідфейки в українських сучасних матеріалах. Предмет дослідження – інструменти виявлення дідфейків на прикладі сучасних матеріалів.

За твердженням А. Кусого, «дідфейк – це найвідоміша форма того, що називають «синтетичними носіями»: зображення, звук та відео, які, здається, були створені традиційними засобами, створюються за допомогою складного програмного забезпечення» [1]. Висвітленню цього явища приділяли увагу науковці Владислав Погорілий, Владислав Білоцький та Станіслав Борис [2], Оксана Брошнівська [3], Ніколя Обрієн [4], Агнешка М. Вальорска [5]. Вони досліджують проблему дезінформації й дідфейків і надають корисні поради суспільству, пояснюють процес створення дідфейків, способи їхнього розпізнавання й наслідки поширення. Крім того, у деяких наукових працях вищезазначених дослідників розглянуто правові аспекти цієї проблеми й запропоновано способи її розв'язання.

Штучний інтелект є ідеальним інструментом для створення й поширення дезінформації завдяки своїм можливостям, зокрема йдеться про генерацію візуального контенту, голосових повідомлень і текстів. За твердженням Ю. Лучко, «можливості штучного інтелекту використовують для створення «підроблених людей» (fake people), тобто створення облич або персонажів, які не існують в реальному житті, за допомогою технологій штучного інтелекту та комп'ютерної графіки. Це може включати створення фотографій, відео або 3D-моделей, які виглядають дуже реалістично, але фактично є створеними комп'ютером» [6, с. 105].

Розглянемо деякі приклади сучасних дідфейків і можливі процеси їхньої ідентифікації. 2021 року в мережі Tik-Tok з'явився фейк, у якому використано обличчя актора Тома Круза [7]. У цьому відеоролику несправжній Том Круз демонструє фокуси. Для виявлення цього дідфейку будемо використовувати

програму Deepware [8]. Scanner.deepware.ai – це вебсайт, який використовує ШІ. Використовуючи його, користувачі можуть сканувати різноманітні типи зображень, відео й отримувати повний аналіз. Програма аналізує за допомогою 4 неймереж: Deepware, Seferbekov, Ensemble й Avatarity.

Подивившись відео із соцмережі Tik-Tok, ми помітили: у міміці Тома Круза є аномалії, що можуть свідчити про потенційне використання технології діпфейку. Після додаткового аналізу в неймережі Deepware були отримані метадані відео, включаючи назву, розмір, тривалість і дату публікації, а також шкалу оцінки наявності діпфейку (рис. 1).



Рис. 1 Результати перевірки відео з Томасом Крузом на виявлення діпфейку.

У результаті стрілка вказала на червону зону та напис *deepfake detected*. Це означає, що відео було створено за допомогою штучного інтелекту й не є автентичним матеріалом. Для додаткової перевірки було зроблено скріншот із відео, який проаналізували за допомогою пошуку зображень у Google. Інтернет-джерела підтвердили, що це відео фальшиве.

Під час повномасштабного вторгнення на територію України країна-агресор почала активно використовувати технологію діпфейку, створюючи образи військових, президента нашої держави для введення в оману українців.

Проаналізуємо відео з ексголовнокомандувачем Збройних сил України Валерієм Залужним. У цьому відео він нібито після новини про свою відставку планує зробити переворот за допомогою військових і негативно висловлюється щодо діяльності В. Зеленського.

Під час перевірки цього відео програмою Deepware ми отримали метадані й інформацію від чотирьох джерел: Deepware, Seferbekov, Ensemble й Avatarity. Кожне з чотирьох джерел офіційно підтвердило, що це відео фальшиве. Avatarity стверджує, що у відео лише 19 % неправдивого матеріалу, три інші джерела вказали 93 % (рис.2).

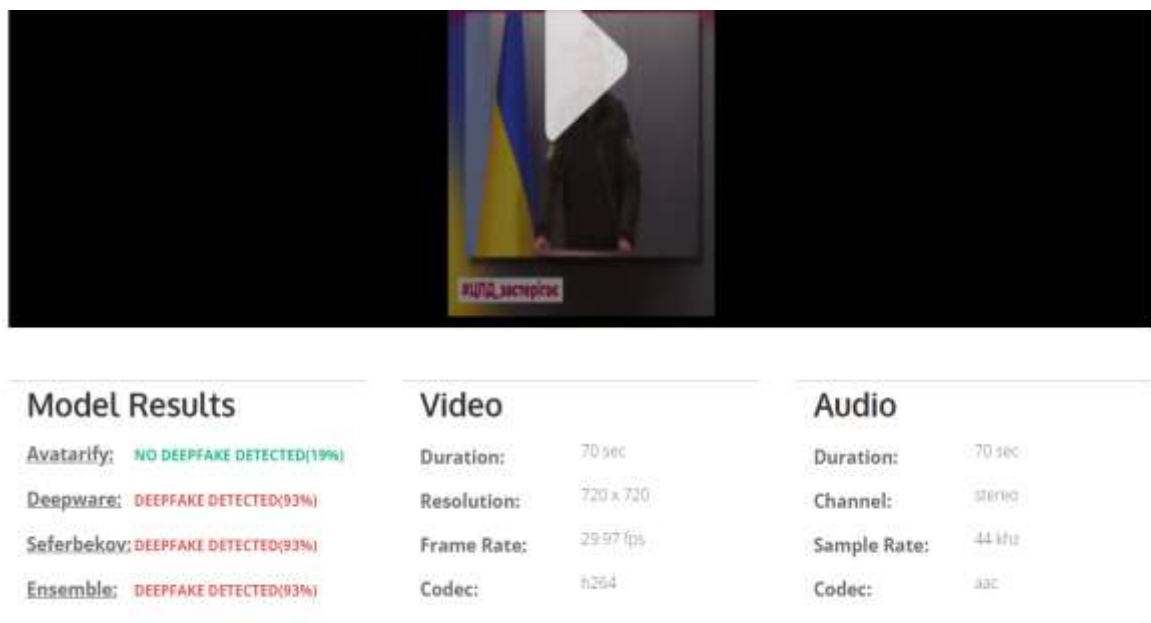


Рис. 2 Результат аналізу відео з Валерієм Залужним на виявлення діпфейку.

Крім того, ми подивилися останні публікації Валерія Залужного й виявили, що не було жодного схожого відео з таким контекстом. Фейкове відео було створено на основі відеовиступу генерала 6 грудня 2023 року з нагоди Дня ЗСУ.

Ми перевірили на правдивість відео з Президентом України. У відео, яке публікували російські пабліки, Володимир Зеленський нібито розповідав, як

важко бути Президентом України, і закликав військових скласти зброю й повернутися до своїх сімей.

По-перше, у цьому відео неважко помітити дуже активні рухи головою Президента України, хоча його тіло не рухалося. По-друге, результати використання програми Deepware підтвердили, що це відео – діпфейк (рис. 3): Deepware, Seferbekov, Ensemble визнають, що це відео є фальшивим (Deepware 99%, Seferbekov 92%, Ensemble 93%), Avatarity зазначає 19% несправжності.

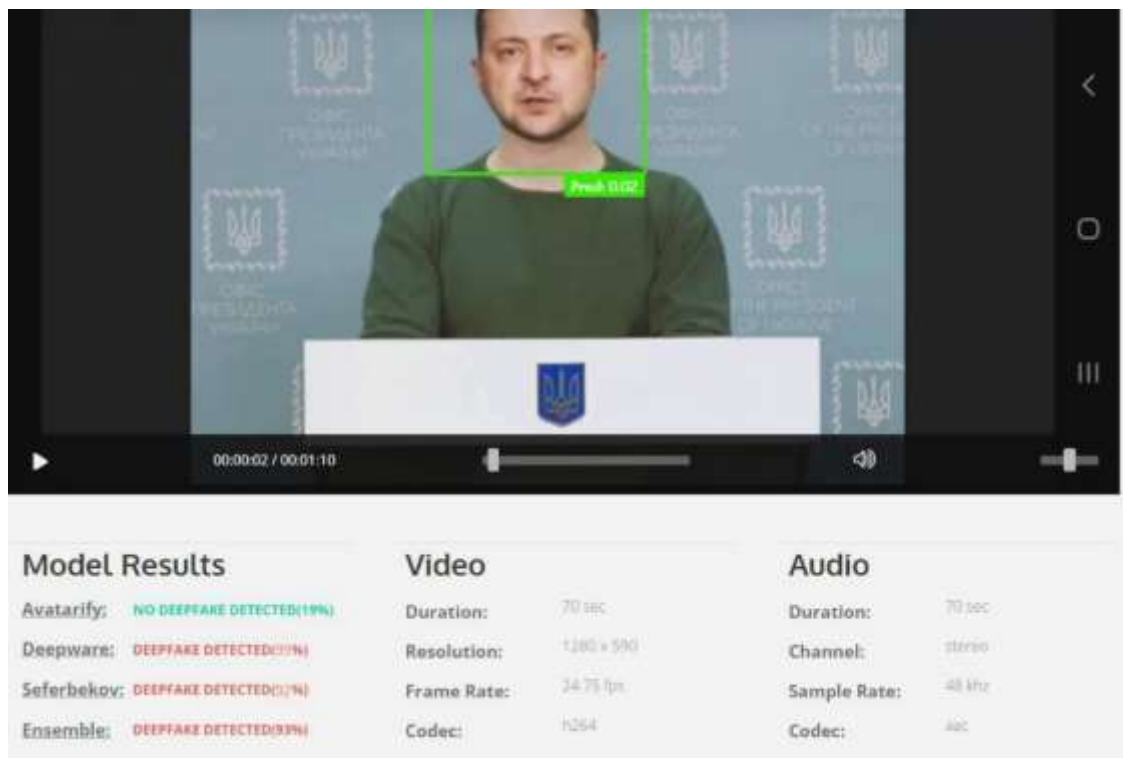


Рис.3 Результати аналізу відео із Президентом України Володимиром Зеленським на виявлення діпфейку.

Хоча використання безкоштовної програмної системи Deepware може допомогти у виявленні використання штучного інтелекту у відеоматеріалах, його ефективність не завжди 100%.

Отже, щоб зробити остаточний висновок про правдивість відео, необхідно проводити його додатковий аналіз, а не використовувати лише один інструмент. Щоб виявляти діпфейки й протидіяти їм, необхідно постійно вдосконалювати технології виявлення. Щоб користувачі могли відрізнити справжню інформацію від фальшивої, важливо підвищувати їхню медіаграмотність.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Кусий А. Діпфеки: як розпізнати та захиститися? URL: <https://netfreedom.org.ua/article/dipfejki-yak-rozpiznati-ta-zahistitisya> (дата звернення 24. 02. 2024)
2. Білоцький В., Погорілий В., Борис С. Чому діпфейк – злочин і як з ним боротись? Пояснення адвоката. URL: <https://ain.ua/ru/2021/07/12/chomu-dipfejk-zlochyn-i-yak-z-nim-borotis-rolyasnennya-advokata> (дата звернення 24. 02. 2024)
3. Брошнівська О. Брехня – нова правда. Що таке діпфейки та як їх розпізнати? URL: <https://vctr.media/ua/yak-rozpiznavatu-dipfeyki-50266/> (дата звернення 24. 02. 2024)
4. O'Brien N. Machine learning for detection of fake news. Diss. Massachusetts Institute of Technology. 2018. 56 с.
5. Агнешка М. Вальорска. Діпфейк та дезінформація : практ. посіб. / Агнешка М. Вальорска ; пер. з нім. В. Олійника. К.: Академія української преси; Центр Вільної Преси, 2020. 36 с.
6. Лучко Ю. І. Роль технологій штучного інтелекту в поширенні та боротьбі з дезінформацією. Протидія дезінформації в умовах російської агресії проти України: виклики і перспективи. 2023. С. 104 – 106.
7. Metaphysic.ai. I love magic! (2021). URL: <https://vm.tiktok.com/ZMM1fWyyN/> (дата звернення 24. 02. 2024)
8. Deepware.ai. URL: <https://scanner.deepware.ai> (дата звернення 24. 02. 2024)